

WWWの情報統合による企業評価システムの構築

Construction of system to evaluate company using information integration from WWW

野路 ひとみ^{*1}

Hitomi NOJI

新美 礼彦^{*2}

Ayahiko NIIMI

小西 修^{*2}

Osamu KONISHI

^{*1}株式会社グッドウィル・エンジニアリング

Goodwill Engineering

^{*2}公立はこだて未来大学 システム情報科学部

Future University-Hakodate, School of Systems Information Science

Recently, to increase the rate of job separation is the one of the most important problem. There main reasons to separate job are complaint about working condition and anxiety among company's possibility. For this problem, it is most important things to prevent the rate of job separation that job seeker before joining the company. In this research, it is proposed the system for job seeker to evaluate company used management, recruiting, and stock quotation on the WWW. And it integrates the information in this system. Finally, it calculate index to evaluate of company and display these results. The index is easy to understand company's information for job seeker who is unfamiliar with financial information. This system takes in the company's information on the Web. For that purpose, this system can provide latest information than other existing systems. By using proposed system, we will have many applications to be able to show inter-company comparison in job description, category of business in the future.

1. はじめに

本研究では、企業の情報を WWW 上より検索し、これまで個々に表されていた情報を統合し、それらの情報を、企業評価を表す指標として加工することで、求職者に対してよりわかりやすい企業情報を提供するシステムを提案する。そして、統合された企業情報を知ることで、就職活動の際に企業と求職者の間に起こるニーズのギャップを解消することを目的とする。本研究で作成したシステムでは、比較的フォーマットが決まっている 3 つの Web サイトに的を絞って情報の抽出を行った。決まった Web サイトを用いることで、同じ情報を得ることができるため、多くの情報を一度に得ることができるだけでなく、企業間の比較を容易に行えるということがわかった。しかしながら、決まった項目がないと正しい結果を表示させることができないという問題点が残った。これらの問題の解決策としては、情報を抽出する Web サイトを増やすこと等が挙げられる。

2. 背景と目的

近年、若年者の離職率が年々高くなっている。リクルートワークス研究所の調査結果によると、転職理由として、勤務条件等の不満が挙げられており、就職前と就職後の企業に対するイメージのギャップが大きいものであると考えられる。[works-i] これらの就業問題に関する不満を解消するためには、就職前に企業の情報を深く知っておく必要があると考えられる。本研究ではそのような情報を知る方法として Web 上にある企業情報より、情報統合を行い、企業の経営情報を指標化したものを提供するシステムを作成することを目的とする。将来的には、就職情報を提供できるサービスとして利用できるようにするこ

とが考えられる。

3. 提案するシステム

構築するシステム内での動作について、以下を提案する(図 1)。このシステムでは、ユーザが指定した Web サイト内にある調べたい企業についての URL を入力することで、その企業についての情報を得ることができる。システム内の構造は、大きく分けて、2 つのシステムで構成されている。

1 つはユーザの入力した企業についての検索を行い、情報を抽出するシステム、もう 1 つはそれらの企業情報を統合するシステムである。

システムは、次の 1 から 5 の流れで動作する。それぞれの動作の位置づけは図 1 の通りである。

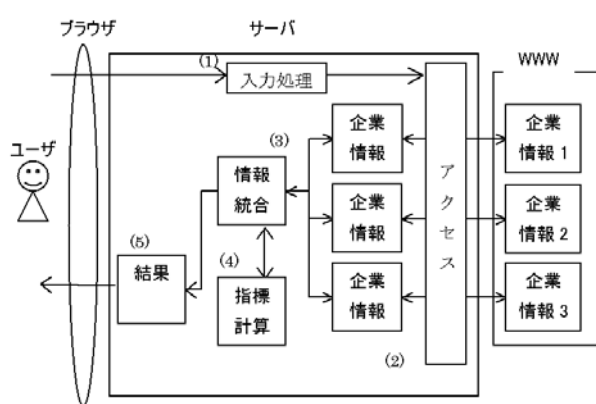


図 1: システム構成図

1. ユーザの入力した URL を読み込み、Web ページを保存する。

連絡先: 〒 041-8655 北海道函館市亀田中野町 116-2

公立はこだて未来大学 システム情報科学部

新美 礼彦

TEL:0138-34-6222 FAX:0138-34-6301

E-mail:niimi@fun.ac.jp

2. 保存された Web ページより情報抽出を行い、XML ファイルに保存する。
3. 2. で抽出された複数の XML ファイルを 1 つに統合する。
4. 3. で統合された XML ファイルより指標計算を行い、3. で作成した XML ファイルに追加する。
5. 4. で作成された XML ファイルより結果を Web ブラウザに表示する。

このシステムの結果として、企業の基本情報や財務状況、採用情報、指標を表示する。基本的な情報だけでなく、第 3 者的に企業を判断できる指標を用いることで、企業についてより詳しい情報を得ることができると考えられる。このシステムでは、Web 上の様々な形式のファイルを扱うことを想定しているため、XML を用いている。また、プログラムでは Java を使用している。Java 言語は、他の言語と異なりユーザのコンピュータの環境に関係なく、プログラムを実行できるため、有用であるため用いた。

これらの動きをシステム内では大きく、情報検索抽出部分と情報統合部分・指標計算部分の 3 つに分類する。情報検索抽出部分は、図 1 では (1) から (2) の部分、情報統合部分は (3) の部分、指標計算部分は (4) に相当するものとして考える。これらの詳しい動作については次項で説明を行う。

3.1 情報検索抽出

システムでは、まず Web ブラウザより入力された情報を読み込み、Web 上のサイトを保存し、その保存した HTML ファイルから必要な情報を検索、抽出する。この際に、Java のパターンマッチのクラスを用いて、HTML のタグより企業情報のテキスト部分の抽出を行う。抽出された各項目の情報は、指定した XML のタグの中に入れる。必要な情報は、NIKKEINET [NIKKEINET], 東洋経済 Web [東洋経済], リクナビ 2006 [リクナビ] の 3 つの Web サイトより抽出する。NIKKEINET からは株価情報、そして、貸借対照表・損益計算書より主な財務情報、リクナビ 2006 からは就職情報、東洋経済 Web からは補足的な就職情報を抽出する。この作業によって、参照したサイトごとに 5 つの XML ファイルが作成される。



図 2: 情報の抽出

3.2 情報統合

ここでは、情報統合では Java プログラムを用いて Web サイトから必要な情報を抽出し、それぞれの Web サイトごとに一つのファイルにまとめる。本研究では、様々な形式の Web サイトから情報を抽出することを考え、多くのファイル形式に対応することが可能な XML を用いた。情報統合の際には、Java プログラムを用いて XML 内のタグごと情報を抽出し、各 Web サイトから情報を抽出して個々に作成した XML ファイルより、一企業の情報として一つの XML ファイルにまとめる。

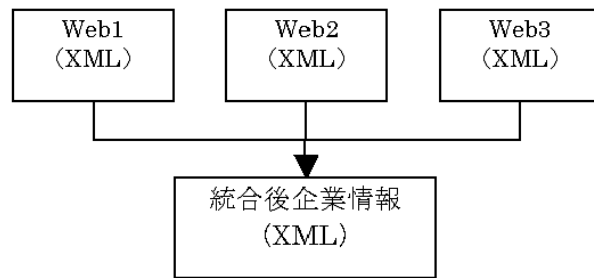


図 3: 情報の統合

3.3 指標計算

企業情報についてより詳しい情報を知るために第 3 者的な位置から判断する材料として、本システムでは指標を用いた企業評価を行う。ここでは 2-1 で抽出された情報を用いて企業評価用の指標計算を行い、企業情報の XML ファイルに追加する。ここで用いる指標は、FUJITSU Cyber Seminar を参考に作成している。[FUJITSU Cyber Seminar] 本システムでは Web から抽出される情報が主に、貸借対照表・損益計算書・株価情報からの情報であるため、そのような情報を用いて算出される指標ということで、上記の指標を参考とした。

1. 収益性...企業の経営において売上に対して利益を示す。
2. バランス性...企業の資産・資本・のバランス性を表す。
3. 効率性...企業の資金が効率よく運営されているかを示す。
4. 株価...一株あたりの価値を示す。

3.4 結果の表示

企業情報は、企業情報とこのような指標計算より作成された指標計算結果を合わせた XML ファイルを HTML 形式に変換し、表示する。この時、計算された指標のほかにも、企業の仕事内容・給料・支社等の情報もテキストの形で表示する。

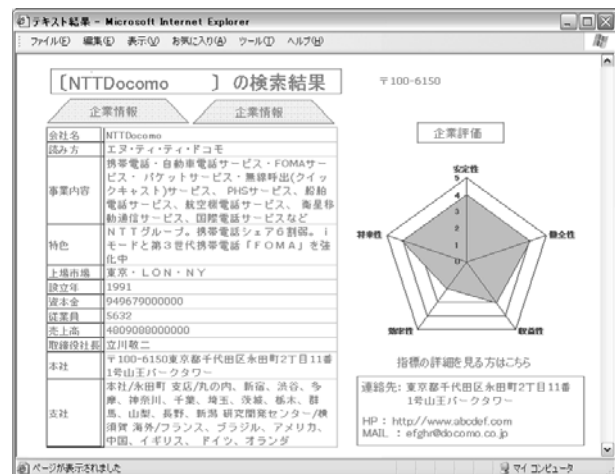


図 4: 結果表示

図 4 のように、企業名や設立年などの指標化に用いることのできない情報についてはテキストの形式で表示を行う。また、指標については例えば、スターチャートのような形の図表で表す。

4. 実験・結果

実験では主に、システム内での情報抽出・情報統合・指標計算の動作実験と、それらの動作実験より、企業評価を提供するシステムとしての動作実験の2種類を行う。それぞれの実験に用いた企業情報は(1) 株式会社 NTT Docomo(2) 東芝ソリューション株式会社(3) 株式会社フジテレビジョンの3社である。

各部分での動作確認では(1) 情報抽出(2) 情報統合(3) 指標計算・結果表示の3種類の実験を行った。結果としては、それぞれの実験において、プログラムは問題なく動作した。しかし、(1)の実験の際に、抽出する情報が欠如しているという問題が起こり、結果表示を行う際にいくつか項目がない部分が出るという問題が生じた。

次に前実験確認した個々の部分の動作確認より、企業評価を提供するシステムとしての動作確認を行う。実験では、前回で用いた企業名と同じ企業についての企業評価を行い、表示結果との比較を行う。この実験は、未だ行っていないが現状では、実験1で使用されたプログラムは動作しており、また、入力されたURLよりWeb上のページを保管することができている。そのため、作成したプログラムを組み合わせ、動作を行うことが実験2の中心となる。実験結果の予想としては、実験1より、システムの動作は問題ないと思われる。しかし、URLを調べなければ企業情報の提供ができない点などからは、ユーザ側としての使いやすさという点では問題があるものと思われる。

5. 考察

実験結果より、本研究で構築したシステムの特徴としては、以下のことが挙げられる。

1. 個々のWebサイトだけでは分からない総合的な情報が得られる
2. 通常のDBを用いるシステムと異なり、Web上の情報を使用することで、常に新しい企業評価を提供できる
3. 指標で表すことで、経営知識がない人でもわかりやすい
4. 参照するWebサイトの項目名やレイアウトが変更されると、正しい情報の抽出が行うことができない

このような特徴より、ユーザにとって、企業を知るための多くの情報を提供することができ、また指標を用いていることで分かりやすい情報を提供することができるということが出来る。このことより、離職者が必要としているような、より詳しくわかりやすい企業情報の提供が可能であると思われる。しかし、システム内の問題として、必要な情報がないと正しい指標として表示されないということや、ユーザの入力する情報が多い、という点において、問題があると考えられる。これらの解決策としては、抽出する情報源を増やすことや、企業検索の一つとして、企業コードを用いた検索を行うことなどが挙げられる。

6. 今後の展開

本研究では、Web上からの情報抽出・情報統合を用いることで、ユーザに対して、一つの企業検索システムからは得られない多くの情報、そして、指標のような第3者的な立場から見た企業の情報など、多方面からみた企業情報を提供することが

できるといえる。これは、貸借対照表や損益計算書などのデータを見て判断するよりも、経営データの見方がわからない人や若年者でも簡単に企業の特徴や傾向を簡単に知ることができるため、求職者にとって非常に有用であること挙げられる。従って、本システムの利用は、企業を深く知ることができ、企業に対する就業前と就業後のイメージのギャップを解消することができるという点で、有用であると考えられる。

一方で、いくつかの問題が挙げられる。その一つとして、著作権問題がある。その点について、現在Web上の情報の総合的な情報を守るための法律は制定されていないものの、Webを利用する上でのモラルとして、それぞれのWebサイトに対して許可を取るなどの対策が考えられる。また、現在、XMLタグの標準化に関しては財務諸表の項目を標準化するタグ言語XBRLが考案されているが、未だ実装は行われていない。[XBRL] しかし、このタグを用いることが可能となれば、本研究で提案したよりも多くの企業情報を、より簡単な方法でWeb上より抽出することが可能となり、本研究で問題となったような必要情報の欠如等の問題に左右されることなく、多くの企業情報の検索・統合した情報の提供が可能となると考えられる。

参考文献

- [works-i] リクルートワークス研究所, <http://www.works-i.com/>
- [NIKKEINET] NIKKEINET, <http://www.toyokeizai.co.jp/index.html>
- [東洋経済] 東洋経済 Web, <http://www.toyokeizai.co.jp/index.html>
- [リクナビ] リクナビ 2006, <http://www.rikunabi2006.com/>
- [FUJITSU Cyber Seminar] FUJITSU Cyber Seminar, <http://glovio.fujitsu.com/cybersmr/>
- [XBRL] XBRL Japan, <http://www.xbrl-jp.org/>